Feature taxonomy:

What type of features do children associate with categories and how do they fare in predicting category judgments?

Farah Mutiasari Djalal

Gert Storms

Eef Ameel

Tom Heyman

University of Leuven

Correspondence address:

Farah Mutiasari Djalal

Faculty of Psychology and Educational Sciences

University of Leuven

Tiensestraat 102 Bus 3711

B-3000 Leuven

Belgium

Tel. 32 16 37 32 12

E-mail: farahmutiasari.djalal@kuleuven.be

Abstract

The present study investigates category intension in school-aged children and adults at two different levels of abstraction (i.e., superordinate and basic level) for two category types (i.e., artefacts and natural kinds). We addressed two critical questions: what kind of features do children and adults generate to define semantic categories and which features predict category membership judgment best at each abstraction level? Overall, participants generated relatively more entity features for natural kinds categories, compared to artefact categories, as well as for basic level categories, compared to superordinate categories. Furthermore, the results showed that older children and adults generated relatively more entity features than younger children. Finally, situation features play the most important role in the prediction of category judgments at both levels of abstraction. Theoretical implications and comparable results from previous studies are described in detail.

Keywords: feature taxonomy; category intension; category learning

1. Introduction

The world is a complex place that consists of an infinite number of different stimuli. Humans try to overcome the problems posed by this diversity by cutting up the environment into a classification structure, a process called categorization. In order to categorize, people need to focus on attributes or features related to particular concepts so that they can simplify the different stimuli into these concepts. Using these features people can decide whether an object belongs to a certain concept or not (e.g., Malt & Johnson, 1992; Vanoverberghe & Storms, 2003). However, the quality of potential features may differ. Some attributes are better suited to define category membership than others. Furthermore, the importance of a feature in defining category membership can differ across age groups (e.g., Mervis, 1987).

Over the years, several methodologies have been used to investigate what kind of features play important roles in delineating categories, both in studies with adult participants and in developmental studies. The techniques varied from studying dictionary definitions (Farah & McClelland, 1991), over providing featural descriptions of to-be-categorized exemplars (Gelman, 1988; Hampton, Storms, Simmons & Heussen, 2009; Keil & Batterman, 1984), to feature generation studies (Vanoverberghe & Storms, 2003). First, we will provide an overview of these studies' main findings, from which we will then derive the specific predictions that are tested in the present study.

1.1. Studies with adult participants

Two general conclusions can be drawn from the array of studies using adult participants. First, perceptual features (e.g., "has wings") tend to be more important in decisions about category endorsement for natural kind categories, while functional features (e.g., "used to work with") are more important to decide on membership in artefact categories (Barr & Caplan, 1987; Barton & Komatsu, 1989; Farah & McClelland, 1991; Medin & Ortony, 1989; Rips, 1989; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Tversky & Hemenway, 1984)[1]. Contrary to this general view, however, Malt and Johnson (1992) found that membership decisions for artefacts were influenced more by perceptual features than by functional information, leading them to claim that functional features alone are not sufficient to determine membership in artefact categories.

Second, level of abstraction also plays an important role in determining the type of features that define category membership. Rosch et al. (1976) found that functional features were generated most frequently for superordinate categories, while perceptual features were generated most frequently at the basic level. The latter finding was taken to mean that the basic level forms the most abstract level at which an object can be perceptually identified while at the superordinate level differences in physical appearance are not as salient as at the basic level. Rosch et al.'s (1976) claims were partly supported by the findings from a feature generation study from Vanoverberghe and Storms (2003), who found that perceptual features are more important for the basic level, while for the superordinate level, perceptual and functional features are equally important.

1.2. Developmental studies

Many studies have been conducted to investigate the developmental shift of category representations based on features. We will highlight some key findings that served as building blocks for the present study. In one of the earliest studies of this kind Mervis and Crisafi (1982) asked children (aged 2- to 5-year-olds) to divide pictured stimuli into categories defined at different hierarchical levels. They found that the acquisition of feature types is influenced by the order in which category levels are acquired. Since the basic level is acquired first, children

---

[1] Note, however, that the distinction between natural kinds and artefacts is not a strict dichotomy, since there are borderline cases, such as cultivated fruits, bred animals, and mineral compounds.

start predominantly with perceptual features, but as they get older they rely more on functional features, which become more important for the superordinate level.

Using a different paradigm, Keil and Batterman (1984) investigated children's early category representations by focusing on characteristic and defining features. They presented school-aged children (kindergartens, second graders, and fourth graders) with two kinds of stories about potential exemplars of a category (e.g., hats, churches). In one kind of story, exemplars were described as being characterized by correct defining features, which were mostly conceptual in nature, but they lacked important characteristic features, which were mostly perceptual. The second sort of stories consisted of the opposite pattern. Keil and Batterman asked children whether the described instance was a true member of the category. The results showed that there is a shift across age groups from using characteristic, mostly perceptual features to defining, mostly non-perceptual features, and children who are in the transitional phase tend to use both types of features. Thus, the shift might denote an increasing ability to understand concepts and to attend to attributes that are not directly apparent.

Keeping basically the same research paradigm, Keil (1989) followed up on this finding by explicitly investigating the shift in different types of categories, namely natural kinds and artefacts. In one of his studies, he investigated the transformation of objects' identities based on changes in the type of applying features. Keil told school-aged children (5 to 11 years) stories in which certain changes were introduced in natural kind and artefact objects. He found that kindergartners relied on appearance more than on function for both artefacts and natural kinds. Keil argued that, since younger children have shallower theories, they tend to rely more on perceptual (a-theoretical) similarities, while older children rely more on function and less on appearance, especially for artefacts.

Gelman (1988), using an inductive interference task, also investigated the role of different types of features in preschoolers and second graders. She taught the children a new fact (i.e., a new feature) about a category exemplar and checked whether they generalized it to other category members. Second graders tended to find the functional features more generalizable than perceptual features for artefacts, while for natural kinds, the opposite pattern was found. For the preschoolers, this distinction was less pronounced than in second graders.

2. Current study

Summarizing, various developmental studies suggest that young children seem to focus on perceptual features. As they get older, they evolve towards a focus on functional features, but this is more pronounced for artefacts. However, even though all of the described papers tried to find out which (kind of) features are important in the categorization process of children, none of the studies addressed the question directly by asking children to *generate* relevant features. The present study tries to fill this gap by having children of different age groups, as well as adults, sum up features that are important to determine membership for a set of natural kinds and artefacts. To the best of our knowledge, this is the first study that allows a systematic comparison between groups of different ages in terms of the kind of features (perceptual versus functional) they *generate*.

We also wanted to investigate which features *predict* category membership judgment best. Furthermore, we examined these two questions using different kinds of categories: natural kinds (e.g., fruit and berries) and artefacts (e.g., vehicles, bicycles, etc); as well as categories defined at different levels of abstraction: superordinate (e.g., clothes, musical instruments, etc.) and basic categories (e.g., trousers, guitars, etc). For these purposes, we gathered feature generation data and category judgment scores from different age groups. Generated features were classified into the four basic categories of Wu and Barsalou's (2009) coding scheme: taxonomic, introspective, entity, and situation features. Taxonomic features refer to the position of the concept in the category taxonomy and includes synonyms, superordinate, subordinate, and coordinate categories as well as specific instances of the concept (e.g., car - automobile; cat - animal). Introspective features refer to the mental state evoked by the concept, such as affects and emotions, evaluations, etc (e.g., apples - I like; smashed car - anger). Entity features are features of a concrete entity, such as external surface features, internal surface features, and external and internal components (e.g., apple - red; watermelon - juicy). Finally, situation features refer to situations in which the concept functions in an event with one or more participants, at some place and time (e.g., shirt - worn; car - transport). Crucial for the purpose of our study is that, for the concrete concepts that we used as stimuli, perceptual features are classified as entity features and functional features are coded as situation features[2].

Regarding the type of features people generate, three hypotheses were derived from the described literature. First, we expect predominantly entity features for natural kinds, while for artefact categories, we expect a more even mix of situation and entity features (Farah & McClelland, 1991; Vanoverberghe & Storms, 2003). Second, based on the findings from Rosch et al. (1976) and Vanoverberghe and Storms (2003), entity features are expected to be generated more frequently for the basic level categories, whereas situation features are expected to be generated more frequently for the superordinate level categories. Third, in line with Keil and Batterman (1984) and Keil (1989), we predict that the youngest children start off mainly with entity features and as children get older, they will rely more on situation features. However, the latter might depend on the category type and the level of abstraction. That is, developmental studies suggest that older children (and adults) will still generate (mainly) entity features for basic level categories (Mandler, 2000; Mervis & Crisafi, 1982) and natural kinds categories (Gelman, 1988; Keil, 1989).

[2] Although Wu and Barsalou's (2009) coding scheme contains further subdivisions, we only use the four basic classifications here.

To examine the value of the entity and situation features in predicting categorization decisions we used the family resemblance model from Rosch and Mervis (1975), which has been shown to relate strongly to category related variables (e.g., Ameel, Malt, & Storms, 2008; Verheyen, De Deyne, Dry, & Storms, 2011). The general idea is that objects are more likely to be considered category members if they possess the most important features for that category. Put differently, the higher an objects' family resemblance score, the higher the likelihood of belonging to the category in question. So besides a feature generation and category judgment task, we also obtained feature applicability judgments in order to calculate family resemblance scores (see the Results section for more details). The critical question is whether the entity-based family resemblance scores predict category judgments better or worse than the situation-based family resemblance scores. More specifically, category type, level of abstraction, and age should have an analogous effect on the type of features that *predict* category membership as on the type of features that people *generate* (see above).

3. Method[3]

3.1. Participants

Children of three different age groups and adults performed the feature generation and the category judgment task. A total of 96 children, consisting of 32 5-year-olds (18 males and 14 females; mean age: 5 years and 6 months), 32 7-year-olds (15 males and 17 females; mean age: 7 years and 6 months), and 32 10-year-olds (22 males and 10 females; mean age: 10 years and 7 months) were recruited from two different schools in Flanders, Belgium. Sixteen adults (8 males and 8 females; mean age: 26 years and 10 months) performed the same tasks. An additional 16 adults (6 males and 10 females, mean age: 30 years and 5 months) volunteered

[3] All the data analyzed in the current paper were collected as part of a large study with a different focus, described in Djalal, Hampton, Storms, & Heyman (manuscript in preparation).

to fill out the feature applicability judgment task. All adult participants were friends and colleagues of the researchers, recruited in Flanders, Belgium.

3.2. Materials

Eight semantic categories were selected in pairs of a basic (low level) and a corresponding superordinate (high level) category belonging to the same semantic domain (i.e., trousers-clothes; berries-fruit; guitars-musical instruments; bicycles-vehicles). A total of 120 colored pictures (11x10 cm), 15 per category, were created for the category judgment and feature applicability judgment tasks. Each set of 15 pictures consisted of ten pictures depicting presumed category members and five portraying related non-members (based on discussions of the selected materials by three of the authors). Fig. 1 displays some of the stimuli. All the stimuli used in this study can be found in Appendix A.

In the feature generation task, participants only received the category labels. Participants' responses were processed using McRae, de Sa, and Seidenberg's (1997) procedure and subsequently used as input for the feature applicability judgment task. The former involved tallying all responses per category, after which synonym features (i.e., features with essentially the same meaning such as *to race* and *to compete*) were merged, adjective-noun combinations (e.g., *heavy iron*) and conjunctive features (e.g., *red and small*) were split up (if they provided different information), redundant quantifiers (e.g., *most of them*) were dropped, and exemplars of the category (e.g., *pear* for the category *fruit*) were removed. The resulting feature lists were combined with the 15 exemplars per category to form feature × exemplar matrices.

Fig. 1. Example stimuli of the categories *clothes*, *fruit*, *musical instruments*, *vehicles*, *trousers*, *berries*, *guitars*, and *bicycles* (from the top left to the bottom right).

3.3. Procedure

Participants were tested individually in one test session. Their first task was a feature generation task, which was followed by a category judgment task. The order of these two tasks was fixed across all participants, except for the additional group of adults who performed only the feature applicability task. For the children, both tasks were conducted in a playful context. Whereas adults performed both tasks for all the studied categories, in order to limit the duration of the task, children performed the tasks for four categories, each belonging to a different domain, two of the high and two of the low level (e.g., clothes, musical instruments, berries, and bicycles). The combination of categories was randomized across participants.

For the children, the feature generation task was presented in the context of helping Mr. Mouse and Mrs. Monkey (presented as stuffed animal versions of a mouse and a monkey), who came from another planet, to understand the meaning of the category terms. By means of an example, the experimenter explained to Mr. Mouse and Mrs. Monkey what furniture was by giving features of furniture (e.g., *it is in the house*, *you can sit on it*, *you can put something on* *top of it*, etc.). The child could begin the actual feature generation task when she understood the task instructions properly. The adult participants were simply given an excel file, consisting of eight different sheets, one for each of the studied categories. They too were asked to imagine they had to explain the terms to someone who did not know its meaning. The adults performed the task individually by writing down the features in an excel file.

The category judgment task was also presented to children in a playful context, where this time children had to help Mr. Mouse and Mrs. Monkey to judge which items belonged to the category. The category name was first mentioned and then the 15 pictures were presented one by one. Children were asked whether the pictured exemplar belonged to the category X (e.g., “is this furniture?”). The adult participants performed the task on a computer. They were given a link to an online survey, where each set of 15 pictures was presented and they made their judgments by clicking on the pictures of items they judged to be members of the category.

For the feature applicability task, participants were presented with a matrix with the 15 pictured exemplars as columns and the (47 to 74, depending on the category) generated features as rows. Participants were asked to indicate whether the exemplars possessed the features, by entering a 1 if the feature applied to the exemplar, or a 0 if not. Each participant was randomly assigned to fill out the matrices for two categories. In total, four participants were assigned to each category.

4. Results

4.1. What kind of features do people generate?

4.1.1. Coding

The responses obtained in the feature generation task were first processed following the procedure outlined in the Materials section. The only difference is that category exemplars were not removed this time since they are in fact taxonomic features. The resulting features were then classified into the four basic categories proposed by Wu and Barsalou (2009): taxonomic, introspective, entity, and situation features. The classification resulted in an agreement of 92%. Disagreements were resolved via discussion.

4.1.2. Data summary

In order to get a general idea of the kind of features that people generated, we first simply collapsed the data across the eight categories and four age groups. On average, entity features were produced the most (*M* = 2.04, *SD* = 2.11), followed by situation features (*M* = 1.91, *SD* = 1.73), taxonomic categories (*M* = 0.64, *SD* = 1.25), and introspective features (*M* = 0.18, *SD* = 0.48). Fig. 2 gives a more detailed overview of the feature distribution broken down by category type, level of abstraction, and age. As can be seen, participants generated relatively few introspective and taxonomic features. This observation and the fact that we had no a priori predictions regarding introspective and taxonomic features prompted us to only focus on entity and situation features henceforth.

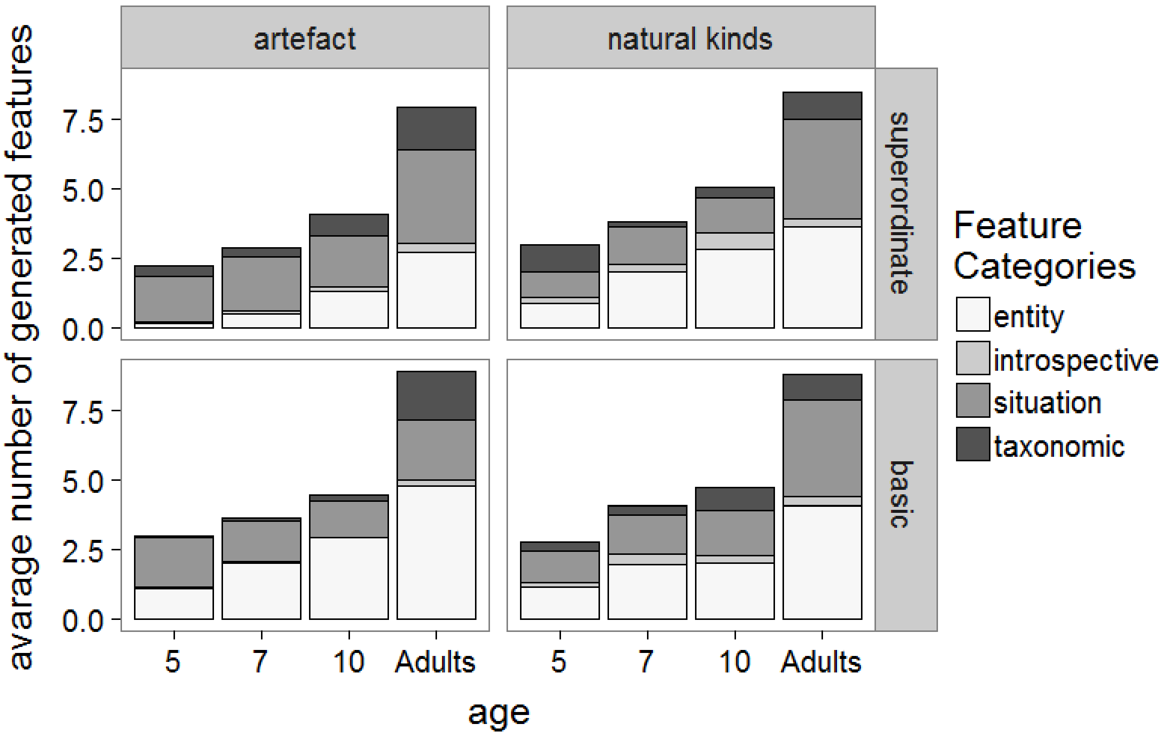


Fig. 2. Feature distribution broken down by category type (artefact vs. natural kinds), level of abstraction (superordinate vs. basic), and age group (5-, 7-, 10-year-olds, vs adults).

4.1.3. Mixed effects analyses

We wanted to test whether the variables category type, level of abstraction, and age were significantly related to the kind of features, entity or situation, people generated. To this end, mixed effects analyses were conducted in which the relative amount of entity versus situation features was predicted. In addition to the fixed effects of category type (natural kinds vs. artefacts), level of abstraction (basic level vs. superordinate), and age (5-, 7-, 10-year-olds, vs adults), we also included random participant effects. Following Barr, Levy, Scheepers, and Tily (2013), a maximal random structure was used (see OSF link: osf.io/ph8uz), for the analysis code and the models that were tested). The analyses were carried out in R (version 3.1.2) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014).

The results revealed main effects of category type ($\chi^2(1) = 17.50$, $p < .001$), level of abstraction ($\chi^2(1) = 67.07$, $p < .001$), and age ($\chi^2(3) = 36.91$, $p < .001$). On average, natural kinds' category labels elicited relatively more entity responses compared to artefacts' category labels. The same is true for basic level categories compared to superordinate categories. Furthermore, as participants get older, they tend to list relatively more entity features (see also Fig. 2). However, these main effects are qualified by two significant interactions (other interactions were not statistically significant). That is, participants produce relatively more entity features for natural kinds compared with artefacts, but this is only true for superordinate categories (category type × level of abstraction: $\chi^2(1) = 32.64$, $p < .001$). Relatedly, participants give relatively fewer entity features for superordinate categories compared with basic level categories, but this only holds for artefact categories. Finally, the tendency of older participants to generate more entity features only manifests itself in the artefact categories (category type × age: $\chi^2(3) = 17.26$, $p < .001$).

4.2. What kind of features predict category judgments?

4.2.1. Data pre-processing

First, to compute the family resemblance scores of every exemplar, the applicability judgments (0 or 1) were summed over the four participants who completed the feature applicability judgment task of each feature across the 15 exemplars, resulting in frequency scores that ranged from 0 to 4 for every item × feature combination. The weights were calculated for every feature by summing the frequency scores of the 10 exemplars that were considered members of the category[4]. The family resemblance score of an exemplar was then calculated by summing, over all features, the product of the feature weight and the corresponding frequency scores. The procedure is illustrated in Appendix B. Family resemblance scores were computed using only those features that were generated by the specific age group at study. In this way, one can learn which type of features generated by a

[4] Note though that including all 15 items in the calculation of the feature weights resulted in very similar results.

particular age group gives the best prediction of category judgments in that age group. So, it is important to recognize that family resemblance scores were computed for both feature types (i.e., entity and situation) separately. The resulting family resemblance scores will be used in a next phase to predict category endorsements. The latter were quantified by scoring each decision as 0 or 1, depending on whether the item was judged a non-member or a member of the category, respectively.

4.2.2. Data summary

To develop some intuitions about what the data look like, we first conducted an exploratory analysis in which item-specific category judgments were obtained by summing across participants per age group. A high category judgment score of an item in a category for a certain age group thus indicates that the item was endorsed often as a category member by that age group, while a low score indicates that the item was usually not judged to belong to the category. Further, in order to investigate which types of features contribute most in predicting category judgments, Spearman rank-order correlations were calculated between category endorsements of each age group and the family resemblance scores based on both types of features, generated by each age group. This was done for all eight categories separately, but in order to simplify matters, we averaged across some categories to focus on the four category level × level of abstraction combinations (see Fig. 3). The results suggest that situation features in general have more predictive value than entity features. To test this more formally, we again conducted mixed effects analyses.

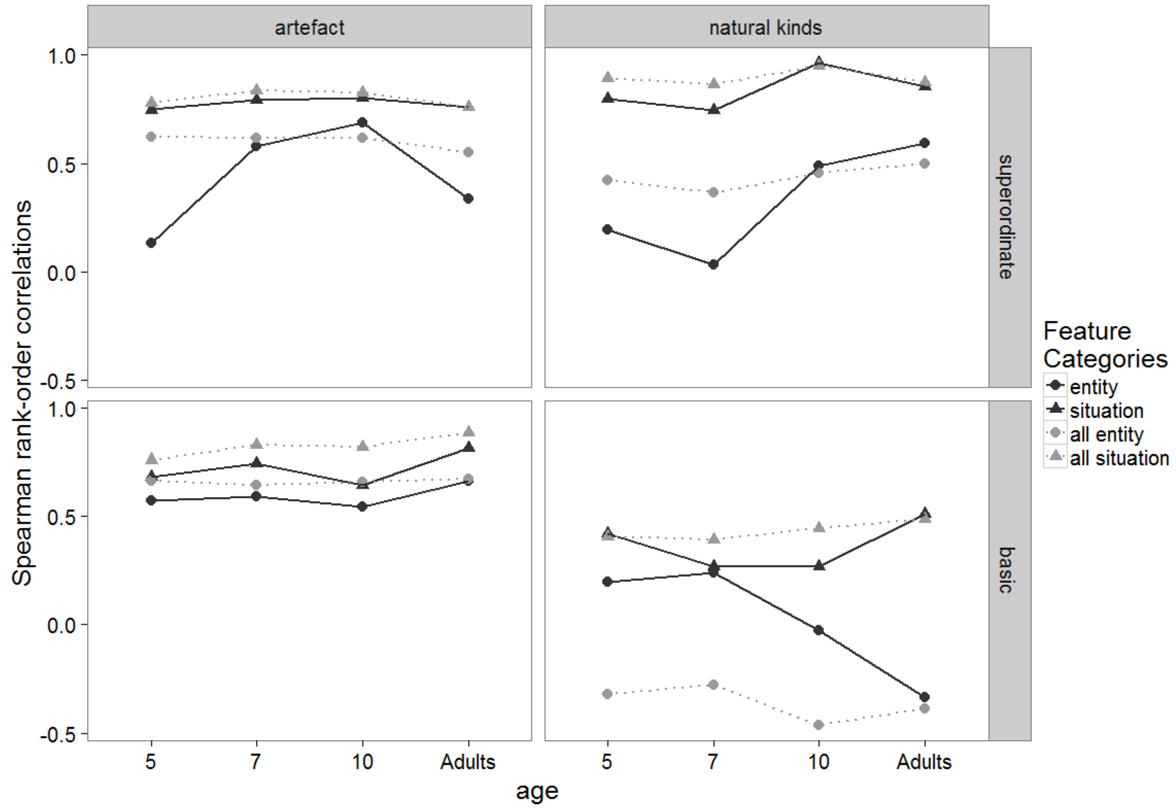


Fig. 3. Spearman rank-order correlations between category judgments and entity-and situation-based family resemblance scores separated per category type (artefact vs. natural kinds), level of abstraction (superordinate vs. basic), and age group (5-, 7-, 10-year-olds, vs adults). As a reference, additional lines (grey dotted lines) were added to show the correlations of category judgments with family resemblance scores based on entity and situation features generated by all four age groups combined.

4.2.3. Mixed effects analyses

The dependent variable of these analyses are the unsummed category judgments (i.e., 0 if an item was considered a non-member, 1 if it was considered a member). In order to reduce model complexity, we conducted separate analyses for all four category type × abstraction level combinations. Hence, the only fixed effects are age, and entity-based and situation-based family resemblance scores. However, when analyzing the artefact data, we also included the variable category (*bicycles*, *guitars*, and *trousers* for the basic level data; *vehicles*, *musical* *instruments*, and *clothes* for the superordinate data). The random structure of the model featured both participant and picture effects, though, because of convergence issues, only random intercepts and random slopes for the main effects (where applicable) were incorporated. Because the latter may yield high Type I error rates when testing interactions (Barr, 2013), we adjusted our alpha level to .01 in those instances. Here, we will only focus on the main effects of the family resemblance variables and interactions involving one or both family resemblance variables.

The situation-based family resemblance score significantly predicted category judgments, but only in the artefact categories ($\chi^2(1) = 4.96$, $p = .026$, for the basic level categories; $\chi^2(1) = 30.98$, $p < .001$, for the superordinate categories). In contrast, the main effect of entity-based family resemblance was never statistically significant. None of the interactions involving any of the family resemblance measures was statistically significant except for the category × entity-based family resemblance interaction in the basic level artefact category data ($\chi^2(2) = 14.80$, $p < .001$). Examining this interaction showed that the entity-based family resemblance did significantly predict category judgments for two out of three categories ($Z = 4.74$, $p < .001$, for *bicycles*; $Z = 2.31$, $p = .02$, for *trousers*).

5. General Discussion

In this study, we investigated category intension, assessed via a feature generation task, and category extension, measured via a category judgment task, in school-aged children and adults. More specifically, we examined a) what kind of features children and adults generate to define semantic categories and b) which features predict category membership judgment best. We will address both questions in two separate sections, which have a similar structure. First, we will summarize the main findings and then we will relate them to previous findings and discuss their implications.

5.1. What kind of features do people generate?

Participants predominately produced entity and situation features, whereas taxonomic and especially introspective features were less prevalent. Furthermore, we found that people generated relatively more entity features for natural kinds categories compared to artefact categories, but only for superordinate categories. Analogously, relatively more entity features were listed for basic level categories than for superordinate categories, but this was only the case for artefact categories. Finally, we found that older participants tend to give relatively more entity features than the young children (5-and 7-year-olds).

The first two hypotheses listed in the introduction are (partly) supported by the data. That is, in line with the literature, it seems that entity features are more salient in basic level categories (Rosch et al., 1976; Vanoverberghe & Storms, 2003) and in natural kinds categories (Farah & McClelland, 1991; Vanoverberghe & Storms, 2003). The only exception seems to be the basic level × natural kinds combination (see Fig. 2), which actually comprises only the category *berries*. One possible explanation is that people, especially children, may not be very familiar with the category *berries*. It is also a fairly heterogeneous category without a clear prototype. Furthermore, family resemblance predictions of category endorsement fail completely in the berries category. This was even the case after including all entity features generated by the four age groups (Fig. 3, light grey dotted lines). It actually suggested that entity features generated by the adults worsened the category membership prediction for the children. Taken together, it is arguably a rather unusual basic level category, which might have obscured the general pattern.

The observed developmental trend (i.e., children begin with situation features and shift more towards entity features as they get older) contradicts with the conclusions emerging from several studies (e.g., Keil & Batterman, 1984; Keil, 1989; Mervis & Crisafi, 1982), in which it was claimed that children shift from relying on perceptual features to both perceptual and functional features, depending on the type of category. This contradiction may be due to differences in the design and focus of the studies. For instance, Keil's (1989) task was very specific to children, as he asked children to judge the importance of different types of features in the transformation of entities. Our study was more straightforward, in the sense that we asked children (and adults) to generate features for different kind of categories. Although this is the traditional way of gathering features (De Deyne et al., 2008; McRae, Cree, Seidenberg, & McNorgan 2005), it is acknowledged that the nature of the task might bias participants to list information that allows them to distinguish the concepts. This would result in generating relatively more features that could separate members from nonmembers, such as situational features, and less features that are shared across objects, such as entity features. However, McRae and colleagues argued that, instead of interpreting this bias as a weakness, it should actually be seen as a strength of this kind of task, since distinguishing (often situational) features play a more important role in object identification in comparison to broad (often entity) features. The idea that functional features are more distinctive than perceptual features is supported by analyses of early-learned noun-feature networks (Hills, Maouene, Maouene, Sheya, & Smith, 2009) and it also resurfaces in the present data. Relative to entity features, situational features tended to have higher distinctiveness values, calculated as the difference between the average applicability judgments for the ten pre-determined members and the average applicability judgments for the five nonmembers (see Table 1).

Table 1

Average distinctiveness scores based on the applicability judgments for each feature category generated by each age group collapsed across the eight categories.

Feature categories	Average distinctiveness score			
	5 yos	7 yos	10 yos	adults
Entity	0.47	0.52	0.62	0.40
Situation	0.65	0.69	0.88	0.82

Furthermore, one might assume that children, even the youngest in our sample, tend to generate functional features more than other types, since they use their experience as their basic knowledge when generating features (Barton & Komatsu, 1989; Gelman, 1988; Barsalou & Wiemer-Hastings, 2005). Indeed, directly asking children to explain the *meaning* of a category label (e.g., *clothes*) might elicit more situational features, because they truly convey the essence of a concept (e.g., *to wear*, *to keep you warm*), especially artefact categories. Adults, on the other hand, may provide some additional, non-crucial, information in the form of entity features (e.g., *it exists in different colors, different styles,…*). This is also in line with Nelson's (1974) *functional core* theory, in which she claimed that children begin with functional features, such as features that relate to possible actors (e.g., mother), to an entity's actions (e.g., throw, pick up, catch), to the location of activity (e.g., living room, playground), and to the effects of action over time (for further discussion, Mandler, 2000). All these features are classified as *situation features* in Wu and Barsalou's (2009) taxonomy.

5.2. What kind of features predict category judgments?

We found that situation features contributed significantly to the prediction of category judgments, but solely for the artefact categories. On the other hand, entity features only had unique predictive value in the categories *bicycles* and *trousers*. In addition, there was no significant interaction with age, suggesting that, as people get older, neither feature type becomes more or less important in determining category extensions.

The finding that situation features predict category judgments for artefacts is in line with several studies in the literature (Barr & Caplan, 1987; Barton & Komatsu, 1989; Rips, 1989) reporting that functional features, such as situation features, are more important than perceptual features (i.e., entity features) for category membership decisions of artefact concepts (but see Malt & Johnson, 1992). The current results also support Barsalou and Wiemer-Hastings' (2005) view that the meaning of a concept is understood and represented against background situations. They claimed that a concept is easier to process when the situation is available. For instance, to understand the meaning of *vehicle*, people do not rely (solely) on the physical entities, but on the setting where they are found (e.g., *roadway*) and the activities performed with them (e.g., *driving*). Knowing the relevant situations where a concept occurs strengthens the knowledge of that concept. Therefore, situation features seem to play an essential part in the representation of concepts. Relatedly, situation features are often defining features, whereas entity features are mostly characteristic features (see also Keil & Batterman, 1984). Therefore, it shouldn't come as a surprise that situation features are better at predicting category judgments, especially given the nature of the feature generation task (i.e., "explain the meaning of X"), which might prompt participants to list more features that discriminate between exemplars and non-exemplars. However, it is remarkable that the entity features have so little predictive value. Based on previous studies and the feature generation data, one would expect entity features to play a (crucial) role in natural kinds categories. Perhaps the low number of natural kinds categories (i.e., two) and the previously mentioned peculiarity of the *berry* category, could explain the unexpected results.

An important finding from our study is that the results suggest there is a disconnection between the type of features people generate and the type of features that delineate category extensions. Even though natural kind categories *elicited* relatively more entity features, there is no significant gain in terms of ability to *predict* category judgments. These findings thus support the notion that there is no perfect link between category intensions and their extensions (Hampton & Passanisi, 2016; Malt, Sloman, Gennari, Shi, & Wang, 1999). Especially (young) children, but even adults (Smith & Medin, 1981), may lack the ability to introspectively decide what features are important to define a category. This was also reflected by the fact that including all entity and situational features generated by the four age groups only improved the category membership prediction modestly (see Fig. 3).

A similar argument applies to the developmental effects: On the one hand, we found a developmental shift from generating situation features to entity features, which seems to partially contradict the aforementioned studies (e.g., Keil & Batterman, 1984; Keil, 1989; Mervis & Crisafi, 1982). However, situation features do play an important role in predicting category judgments across all age groups. Put differently, despite the fact that people seem to add (mostly entity) features as they get older, the feature-based prediction of category judgments does not improve significantly. This would again suggest that there is a discrepancy between “what they generate” (the intensions) and “what they actually use” in judging category membership (the extensions).

Acknowledgments

FMD gathered and analyzed the data presented in this paper, and participated in the writing of the manuscript. TH contributed in the writing of the manuscript and analyzed the data. All four authors discussed the findings thoroughly, read, and approved the final version of the manuscript. The research described in the manuscript was sponsored by grants DBOF/12/010 and OT/10/024 from the Leuven Research Council. We are grateful to Lieselot Verlooy, Liesbeth Berckmans, and Sarah Casaer for assisting in the data collection. TH is a postdoctoral fellow of the Research Foundation-Flanders (FWO-Vlaanderen). Correspondence should be addressed to FMD, Department of Psychology, University of Leuven, Belgium. Email address: Farahmutiasari.Djalal@kuleuven.be.

References

Ameel, E., Malt, B., & Storms, G. (2008). Object naming and later lexical development: From baby bottle to beer bottle. *Journal of Memory and Language, 58*(2), 262-285. doi: 10.1016/j.jml.2007.01.006

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4*, 328. doi: 10.3389/fpsyg.2013.00328

Barr, R. A., & Caplan, L. J. (1987). Category representations and their implications for category structure. *Memory & Cognition, 15*(5), 397-418. doi: 10.3758/BF03197730

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. doi: 10.1016/j.jml.2012.11.001

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. Grounding cognition. In D. Pecher & R. A. Zwaan (Eds.), *The role of perception and action in memory, language, and thought*, (p. 129-163). New York: Cambridge University Press

Barton, M. E., & Komatsu, L. K. (1989). Defining features of natural kinds and artefacts. *Journal of Psycholinguistic Research, 18*(5), 433-447. doi: 10.1007/BF01067309

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-7). Retrieved from: http://CRAN.R-project.org/package=lme4

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M. J., Voorspoels, W., & Storms, G. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods*, *40*(4), 1030-1048. doi: 10.3758/BRM.40.4.1030

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, *120*(4), 339. doi: 10.1037/0096-3445.120.4.339

Gelman, S. A. (1988). The development of induction within natural kind and artefact categories. *Cognitive Psychology*, *20*(1), 65-95. doi: 10.1016/0010-0285(88)90025-4

Hampton, J. A., & Passanisi, A. (2016). When intensions do not map onto extensions: Individual differences in conceptualization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(4), 505-523. doi: 10.1037/xlm0000198

Hampton, J. A., Storms, G., Simmons, C., & Heussen, D. (2009). Feature integration in natural language concepts. *Memory & Cognition*, *37*(8), 1150-1163. doi: 10.3758/MC.37.8.1150

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Categorical structure among shared features in networks of early-learned nouns. *Cognition*, *112*(3), 381-396. doi: 10.1016/j.cognition.2009.06.002

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior, 23*(2), 221-236. doi: 10.1016/S0022-5371(84)90148-8

Nelson, K. (1974). Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review, 81*(4), 267-285. doi: 10.1037/h0036592

Malt, B. C., & Johnson, E. C. (1992). Do artefact concepts have cores? *Journal of Memory and Language, 31*(2), 195-217. doi:10.1016/0749-596X(92)90011-L

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*(2), 230-262. doi: 10.1006/jmla.1998.2593

Mandler, J. M., (2000) Perceptual and conceptual processes in infancy. *Journal of Cognition and Development, 1*(1), 3-36, doi: 10.1207/S15327647JCD0101N_2

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods, 37*(4), 547-559. doi:10.3758/BF03192726

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126*(2), 99-130. doi: 10.1037/0096-3445.126.2.99

Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York: Cambridge University Press

Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), *Concepts and conceptual development: ecological and intellectual factors in categorization* (p. 201-233). Cambridge University Press

Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development*, *53*(1), 258-266. doi: 10.2307/1129660

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. New York: Cambridge University Press

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573-605. doi: 10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382-439. doi:10.1016/0010-0285(76)90013-X

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, *113*(2), 169-193. doi: 10.1037/0096-3445.113.2.169

Vanoverberghe, V., & Storms, G. (2003). Feature importance in feature generation and typicality rating. *European Journal of Cognitive Psychology*, *15*(1), 1-18. doi: 10.1080/09541440303600

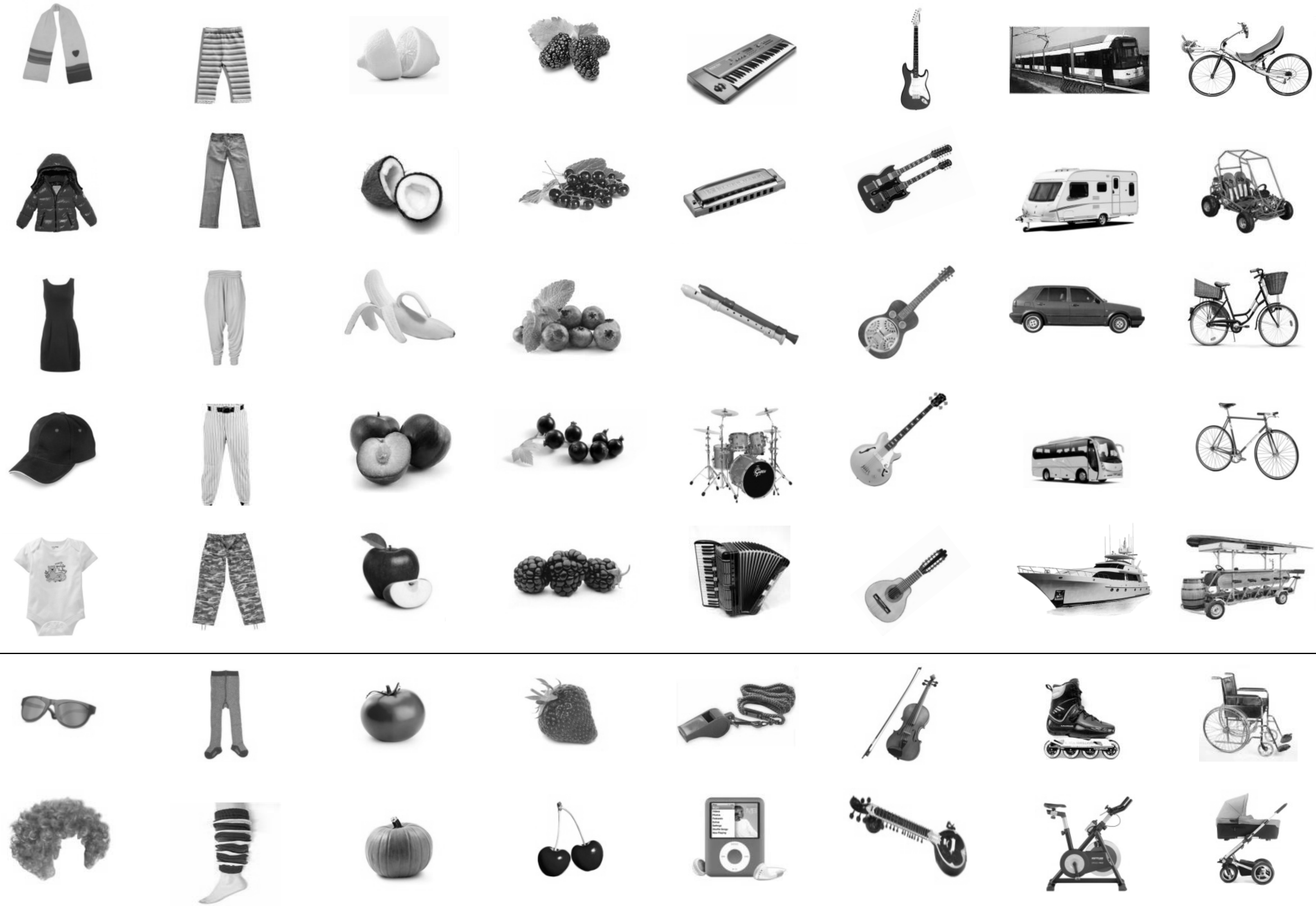
Verheyen, S., De Deyne, S., Dry, M. J., & Storms, G. (2011). Uncovering contrast categories in categorization with a probabilistic threshold model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1515-1531. doi: 10.1037/a0024

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica, 132*(2), 173-189. doi: 10.1016/j.actpsy.2009.02.002

Appendices

Appendix A. Stimulus Set

	Clothes	Trousers	Fruit	Berries	Musical instruments	Guitars	Vehicles	Bicycles
								
								
Members								
								
								



Non-members

Appendix B. Illustration of the procedure to calculate the family resemblance scores

Step 1: Calculating the *feature weights*

All exemplars of the category fruit

						Feature weights
It contains vitamins	4	4	4	....	4	**40**
It has seeds	4	0	0	....	3	**29**
....	....	....	....	....	....	**....**
It exists in different sizes	1	3	3	....	2	**29**
It is to eat	4	+ 4	+ 4	...	2	**40**
It grows on a tree	4	4	4	....	1	**33**
....	....	....	....	....	....	**....**
You can peel it	4	3	4	....	3	**39**

Entity features generated by a specific age group (e.g., 7 year-olds)

Situational features generated by a specific age group (e.g., 7 year-olds)

Frequency scores: the number of participants who judged *bell pepper* to possess the feature 'it contains vitamins'

Feature weights: the row total for the feature 'it is to eat' across the 10 members

Step 2: Calculating the family resemblance scores

All exemplars of the category fruit

It contains vitamins	(4x40) 160	(4x40) 160	160 +	....	160
It has seeds	(4x29) 116	(0x29) 0	0 +	....	87
....	....	....	....	....	....
It exists in different sizes	(1x29) 29	(3x29) 87	87 =	....	58
Family resemblance	**658**	**583**	**626**	**....**	**820**
It is to eat	160	160	160 +	....	160
It grows on a tree	132	132	132 +	....	33
....	....	....	....	....	....
You can peel it	156	171	156 =	....	171
Family resemblance	**938**	**881**	**1012**	**....**	**797**

Family resemblance scores of entity features: the sum of the weighted feature scores of entity features generated by a specific age group (e.g., 7 year-olds)

Family resemblance scores of situational features: the sum of the weighted feature scores of situational features generated by a specific age group (e.g., 7 year-olds)

Weighted feature score: *feature weight* of the feature 'it contains vitamins' multiplied by the score of *bell pepper*: 4x40